引入迷思概念的关键行动编码及其在过程数据诊断分类分析中的应用

高方方　陈琦鹏　詹沛达*

(浙江师范大学心理学院, 金华 321004)

摘　要　迷思概念是指基于个人经验构建的对一些对象、事件或观点的错误理解。相比于识别问题解决技能的缺失，额外识别迷思概念有助于进一步明确学生出现错误的原因。关键行动编码是将传统心理测量模型迁移应用至分析行动序列等过程数据的必要手段。针对现有关键行动编码无法精细识别导致错误问题解决的原因的局限，本研究提出一种引入迷思概念的关键行动编码方式，以实现同时考虑问题解决技能和迷思概念对问题解决过程的影响。通过一则实证研究对比探究不引入和引入迷思概念的两种关键行动编码在过程数据诊断分类分析中的表现；结果主要发现引入迷思概念可实现对被试的更精细化分类，有助于进一步识别导致问题解决成败的具体原因。

关键词　认知诊断; 过程数据; 问题解决; 迷思概念; 行动序列

1　引言

随着计算机软硬件水平的提高，计算机(网络)化测评逐渐成为当前心理与教育测量的主要形式。相比于传统纸笔测验，计算机化测评可基于日志文件抓取被试解决问题时的结果数据(outcome data)和过程数据(process data)。其中，结果数据是指被试完成任务后得到的可反映其任务解决效果的总结性数据，如作答精度和作答时间[1]；而过程数据是指带有时间戳的能够反映被试解决问题过程的人际(机)交互数据(Hao et al., 2015; 李美娟　等, 2020; 刘红云　等, 2023; 刘耀辉　等, 2022)，如问题解决时呈现的操作历程(i.e., 行动序列)或对话文本信息。分析过程数据有助于了解被试的问题解决过程、探究被试的问题解决策略；同时，也使得研究重点从探究“结果是什么”转变为探究“结果是如何产生的”(Greiff et al., 2016)。当然，如附录图 S1.1 所示，过程数据的非标准化结构[2]也对现有的针对有标准化结构的数据的分析方法提出了挑战，如项目反应理论模型、诊断分类模型(diagnostic classification model, DCM)和结构方程模型等传统心理测量模型均难以直接用于分析过程数据。

根据数据分析逻辑，可将现有的过程数据分析方法归为两类：数据挖掘法和统计建模法(Bergner & von Davier, 2019; Mislevy, 2019; 刘耀辉　等, 2022)。前者是数据驱动“自下而上”的探索性方法，旨在从大量数据中通过算法来揭示有意义的新关系、新趋势和新模式，主要包括随机森林、聚类分析、支持向量机等有监督或无监督学习算法(e.g., Chen et al., 2023; Hao et al., 2016; He et al., 2021; Qiao & Jiao,

* 通讯作者：詹沛达，E-mail: pdzhan@gmail.com

[1] 部分已有研究将题目作答时间归入过程数据，但实际上该数据类型是在被试完成任务后才获取的总结性数据；因此，将其归入结果数据比将其归入过程数据更恰当(Fu et al., 2023)。

[2] 标准化数据结构是指数据具有 $N \times I$ 的矩阵结构，其中 N 表示被试总数，I 表示题目总数 (詹沛达, 2022)。

2018; 孙鑫　等, 2018)；而后者是理论驱动“自上而下”的验证性方法，一般基于理论假设构建有关问题解决中认知过程的函数模型，主要包括隐马尔可夫建模、多水平建模和经典心理测量模型的迁移应用等(e.g., Fu et al., 2023; Li et al., 2023; Liu et al., 2018; Han et al., 2022; Xiao & Liu, 2023; Zhan & Qiao, 2022; 付颜斌　等, 2023)。相较于数据挖掘法而言，统计建模法遵循潜在特质决定外显行为的心理学假设，其分析结果具有更高的可解释性，能更明确指出被试在解决问题过程中存在的有关认知过程或知识技能上的不足。

为构建潜在特质与外显行为之间的联接或将适用于标准化结构数据的心理测量模型迁移应用于过程数据分析，统计建模法通常需要对被试的过程数据进行关键行动(操作)编码(key-action encoding)，即判断每一名被试的过程数据中是否包含解决问题所必需的关键行动，并进行编码(如，1 表示“包含”，0 表示“不包含”)。比如，Liu 等人(2018)和李美娟等人(2020)对被试行动序列进行编码时，将被试呈现正确解决问题的行动和未呈现错误解决问题的行动均编码为 1，其他编码为 0。Zhan 和 Qiao (2022)将正确解决问题的关键行动序列拆解为包含长度不同的单一行动或相邻行动序列的虚拟题目(phantom item)，当被试呈现的行动序列包含上述单一行动或相邻行动序列时，其在虚拟题目上的作答被编码为 1，否则为 0。目前关键行动编码的应用已不限于过程数据分析，也被应用于概念图等表现性评价分析之中(e.g., Xia et al., 2023)。纵观已有采用关键行动编码的研究可发现它们有一个共同局限，即编码时仅考虑问题解决能力与技能对行动的影响，忽略了迷思概念(misconception)对问题解决行动的影响。

迷思概念是指基于个人经验构建的对一些对象、事件或观点的错误理解(Martin et al., 2001)，包括非科学信仰、先入为主的理解、天真理论或概念误解等。且被试一旦形成迷思概念就很难被改变(Eggen & Kauchak, 2004)，会对学习和问题解决行为产生持续性影响。通常，在问题解决中，迷思概念会导致错误作答行为(e.g., Bradshaw & Templin, 2014; Kuo et al., 2016; Levy, 2019; Ma et al., 2024)。已有研究指出，相比于识别被试缺失的正确概念与技能，额外识别被试的迷思概念有助于进一步明确学生出现错误的原因，进而实施有针对性的矫正干预(Kuo et al., 2016; Kuo et al., 2018)。然而，已有关于迷思概念的测量研究大多局限于对结果数据的分析。Levy (2019)将迷思概念引入问题解决过程数据分析，探究了迷思概念与典型行为指标的对应关系；但实际上该研究并未深入过程数据本身，而是从过程数据中提取了一些学生解决问题的总结性行为表现，比如，解题方法是否标准(“Standard Solution”和“Incomplete Solution”)以及是否出现某种错误解题方法(如，“Saw As One Unit”和“Everything In Order”)等。

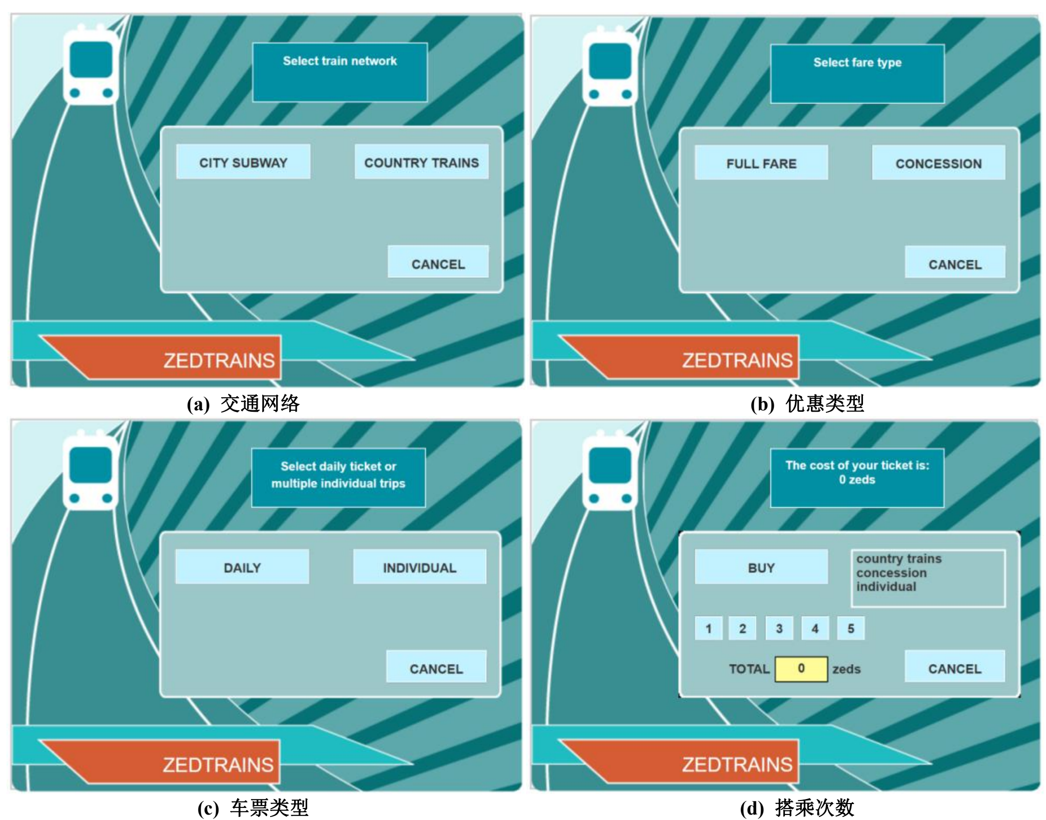
综上，本研究旨在提出一种引入迷思概念的关键行动编码并探究其在过程数据诊断分类分析中的应用，新编码方式可视为对 Zhan 和 Qiao (2022)的编码方式的拓广。

2　方法

2.1　问题解决任务

与 Han 等人(2022)和 Zhan 和 Qiao (2022)一致，本文以 PISA 2012 计算机化问题解决测验 TICKETS (CP038Q02)为例阐述新编码方式及数据分析过程。该任务要求被试购买两趟单程的全价乡村火车票(“Buy a full fare, country train ticket with two individual trips”)，完成任务得 1 分，反之得 0 分。该题目需要被试相继在图 1 所示的页面中进行选择(依次为(a)→(b)→(c)→(d))，即先后在“交通网络”、“优惠类型”、“车票类型”以及“搭乘次数”这四个解题阶段上进行选择；需要注意的是，(1)四个任务界面是有顺序的，被试必须从头做到尾，在点击“BUY”之前没有退出选项；(2)在点击“BUY”之前，无论在哪个页面点击“CANCEL”，都将回到初始界面“交通网络”；(3)只有在“车票类型”环节选择“INDIVIDUAL”后才有在“搭乘次数”环节选择的权利，且此时被试可以漏选搭乘次数直接点击“BUY”结束作答。该任务中，正确解决问题的最优行动序列为 COUNTRY TRAINS→FULL FARE→INDIVIDUAL→2→BUY。

如附录 S1.1 所示，系统记录了每一个学生在解决该任务时每一个阶段的选择，根据“event_value”列中的信息，可以整理出每一个学生的行动序列。当被试呈现的行动序列中包含最优行动序列时，被试将会正确解决该任务；反之，当被试呈现的行动序列中不包含或部分包含最优行动序列时，被试会错误解决该任务。因此，该任务的原始得分仅能将被试分为两类：正确解答组和未正确解答组；无法明确指出被试无法正确解答该任务的原因。



(a) 交通网络 (b) 优惠类型

(c) 车票类型 (d) 搭乘次数

图 1. PISA2012 问题解决测验 TICKETS (CP038Q02)任务操作界面

2.2 被试

本研究的被试选自美国(USA)、新加坡(SGP)、澳大利亚(AUS)和土耳其(TUR)四个国家的 3769 名被

试；删除 4 名基本信息不完善(缺失身份 ID 或学校 ID)以及 218 名中途放弃(没有点击“BUY”)的被试后，剩余有效被试 3547 名。

2.3 引入迷思概念的关键行动编码

2.3.1 行动序列编码

首先，对行动序列进行编码。我们将被试的所有行动或所处的问题状态(problem state)分为三类：起始状态、中间状态和目标状态。首先，起始状态为未进行任何操作时的初始状态(即任务的起始页面)，在 log-file 中被记录为“event”列中的“START_ITEM”。其次，由于被试点击“BUY”后便会结束该任务，所以目标状态为“BUY”。由于每名完成作答的被试都有相同的起始状态和目标状态，无法提供测量信息，本研究不对这两个状态进行编码；使用不同字母对中间状态进行编码，如图 2 所示。

图 2 中，粗实线连接最优行动序列(ABCD)[3]中的正确状态转移，包含 AB、BC 和 CD；细实线连接最优行动序列外的错误状态转移，包含诸如 FG、GH 和 GC 等。由于该任务允许在任一阶段通过点击“CANCEL”来返回起始状态；因此，被试所呈现的行动序列会同时包含正确状态转移和错误状态转移。以附录图 S1.2(a)中被试为例，根据上述编码规则，可将该被试的问题解决行动序列表示为：F → B → “CANCEL” → A → B → C → D → D。进一步，由于“CANCEL”会返回初始状态并在 A 和 F 之间进行选择，所以可以删除“CANCEL”(附录图 S1.2(b))。另外，由于被试可以重复点击某按键，会出现多个连续重复编码的情况；参考 Zhan 和 Qiao (2022)，我们将类似重复问题状态均缩减为 1 个问题状态(附录图 S1.2(c))。最终，该被试的行动序列为 FBABCD，同时包含了正确状态转移和错误状态转移。

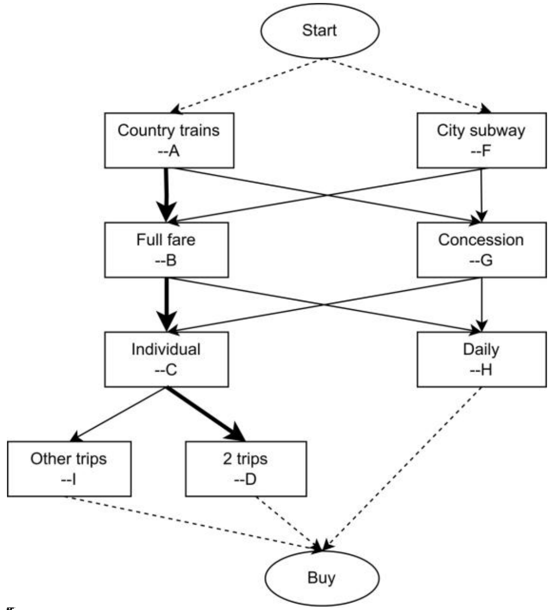


图 2. PISA2012 问题解决测验 TICKETS (CP038Q02)任务的问题状态编码图. 注: “Other trips”代表被试在“搭乘次数”环节选择除 2 张票之外的车票数选项，即 0/1/3/4/5 张票的统称.

---

[3] 除特殊强调或表达需要，本文省略了各状态转移间的箭头符号，即“AB”等同于“A→B”。

本研究使用 N-gram 特征提取，将被试的行动序列拆解为不同长度的字节片段序列(Zhan & Qiao, 2022)。表 1 呈现了该任务下从一元到四元字节片段序列对应的所有行动序列。此时，可以将被试所呈现的任一行动序列视为表 1 中不同元字节片段序列的组合。比如，可将附录图 S1.2 中被试所呈现的行动序列(FBABCD)视为 5 个二元字节片段序列(FB、BA、AB、BC 和 CD)的或 2 个二元字节片段序列(FB、BA)和 1 个四元字节片段序列(ABCD)的组合。总之，经过 N-gram 特征提取，理论上可建构包含所有可能行动序列的行动序列空间(action sequence space)。

表 1. PISA2012 问题解决测验 TICKETS (CP038Q02)任务中 N-gram 字节片段序列.

N-gram	数量	字节片段序列
Uni-gram	8	A、B、C、D、F、G、H、I
Bi-gram	10	AB、AG、BC、BH、CI、CD、FB、FG、GC、GH
Tri-gram	12	ABC、ABH、AGC、**AGH**、BCI、BCD、FBC、FBH、FGC、FGH、**GCI**、**GCD**
Quad-gram	8	**ABCI**、ABCD、**AGCI**、**AGCD**、**FBCI**、**FBCD**、**FGCI**、**FGCD**

注：粗体表示因呈现频率小于 5%或大于 95%而在后续分析中被删除的序列.

2.3.2 Q 矩阵构建

Q 矩阵(Tasuoka, 1983)是诊断分类分析的必要元素。在 DCM 中，Q 矩阵描述潜在属性和题目之间的对应关系，其元素 $q_{ik} = 1$ 表示题目 i ($i = 1, 2, ..., I$)考查了属性 k ($k = 1, 2, ..., K$)，反之，$q_{ik} = 0$。参考 Zhan 和 Qiao (2022)的设定，将行动序列空间中各字节片段视为虚拟题目，在将呈现正确状态转移所需的问题解决技能视为潜在属性的同时还将呈现错误状态转移所需的迷思概念也视为潜在属性；进而，假设仅当被试掌握了各行动序列所需的潜在属性后才能呈现该行动序列。此时，$q_{ik} = 1$ 表示呈现行动序列 i 需要被试掌握潜在属性 k，反之，$q_{ik} = 0$。

(1) 潜在属性界定 鉴于 TICKETS (CP038Q02)任务具有清晰的问题状态转移结构，根据该问题解决任务的题目要求、评分规则和目标认知过程(探索与理解(exploring and understanding)) (OECD, 2014)，可总结出被试需要理解 4 个解题需求才能正确解答该任务(即呈现最优行动序列 ABCD)，分别是(a1)理解需要购买郊区火车票；(a2)理解需要购买全价票；(a3)理解需要购买次票；(a4)理解需要购买 2 张票。同时，被试可能因为具有一些迷思概念而导致其呈现出最优行动序列外的其他行动序列，包括(m1)不理解需要购买郊区火车票；(m2)不理解需要购买全价票；(m3)不理解需要购买次票；(m4)不理解需要购买 2 张票。表 2 列出了 4 个问题解决技能和 4 个迷思概念。

(2) 虚拟题目扩展 首先，遵循 Zhan 和 Qiao (2022)的做法，将表 1 中所有 N-gram 字节片段对应的行动序列视为初选虚拟题目。其次，根据被试作答各初选虚拟题目的人数，删除作答人数少于总人数 5%和大于总人数 95%的(即无法为大多数被试提供信息)的虚拟题目(表 1 中粗体示意的行动序列；所有行动序列的频次图见附录图 S2.1)。然后，考虑到 Q 矩阵完备性和模型可识别性的要求(丁树良 等, 2010; Gu & Xu, 2020)，Q 矩阵至少保留一个可达矩阵。最终，形成包含 28 题和 8 个属性的 Q 矩阵，如表 3

所示。

2.3.3 虚拟题目作答数据编码

针对 28 道虚拟题目，当被试的行动序列包含特定虚拟题目所对应的行动序列时，则该被试“正确作答”该题目；反之，当被试的行动序列没包含特定虚拟题目所对应的行动序列时，则该被试“错误作答”该题目。以附录图 S1.2 中被试所呈现的行动序列(FBABCD)为例，他/她的作答数据向量可被编码为(1111100010100110001000100001)’。

表 2. PISA2012 问题解决测验 TICKETS (CP038Q02)任务的潜在属性及其对应的典型行动序列.

潜在属性		属性含义
问题解决技能	a1	理解需要购买郊区火车票
	a2	理解需要购买全价票
	a3	理解需要购买次票
	a4	理解需要购买 2 张票
迷思概念	m1	不理解需要购买郊区火车票
	m2	不理解需要购买全价票
	m3	不理解需要购买次票
	m4	不理解需要购买 2 张票

表 3. PISA2012 问题解决测验 TICKETS (CP038Q02)任务的 Q 矩阵.

题号	题目	问题解决技能				迷思概念			
		a1	a2	a3	a4	m1	m2	m3	m4
1	A	1	0	0	0	0	0	0	0
2	B	0	1	0	0	0	0	0	0
3	C	0	0	1	0	0	0	0	0
4	D	0	0	0	1	0	0	0	0
5	F	0	0	0	0	1	0	0	0
6	G	0	0	0	0	0	1	0	0
7	H	0	0	0	0	0	0	1	0
8	I	0	0	0	0	0	0	0	1
9	AB	1	1	0	0	0	0	0	0
10	AG	1	0	0	0	0	1	0	0
11	BC	0	1	1	0	0	0	0	0
12	BH	0	1	0	0	0	0	1	0
13	CI	0	0	1	0	0	0	0	1
14	CD	0	0	1	1	0	0	0	0
15	FB	0	1	0	0	1	0	0	0
16	FG	0	0	0	0	1	1	0	0
17	GC	0	0	1	0	0	1	0	0
18	GH	0	0	0	0	0	1	1	0
19	ABC	1	1	1	0	0	0	0	0
20	ABH	1	1	0	0	0	0	1	0
21	AGC	1	0	1	0	0	1	0	0
22	BCI	0	1	1	0	0	0	0	1
23	BCD	0	1	1	1	0	0	0	0
24	FBC	0	1	1	0	1	0	0	0
25	FBH	0	1	0	0	1	0	1	0
26	FGC	0	0	1	0	1	1	0	0
27	FGH	0	0	0	0	1	1	1	0
28	ABCD	1	1	1	1	0	0	0	0

2.4 分析

本研究提出的行动序列编码也可以直接利用已有的 DCM 参数估计软件进行数据分析，比如 R 软件中的 GDINA 包(Ma & de la Torre, 2020)等，一定程度上减少了方法的实践应用难度。本文选用 GDINA 包进行诊断分类分析，具体的使用教程可参阅 GDINA 包使用手册或 Shi 等人(2021)的教程。

由于不确定 8 个属性对行动序列的影响机理，我们同时使用 GDINA (de la Torre, 2011)、DINA (Junker & Sijtsma, 2001)、DINO (Templin & Henson, 2006)和 ACDM (de la Torre, 2011)进行数据分析，然后根据模型-数据拟合指标选取最合适的模型并解读其结果。其中，DINA 模型假设属性之间满足连接缩合规则，即被试必须掌握题目所要求的所有属性，其理想正确作答概率才为 1；DINO 模型假设属性之间满足分离缩合规则，即被试只要掌握题目所要求的任一属性，其理想正确作答概率就为 1；ACDM 假设属性之间满足补偿缩合规则，被试的理想正确作答概率随掌握的属性的数量的增加而增加；GDINA 是一个饱和模型，不局限于特定的缩合规则，具有理论上的普适性。

模型-数据相对拟合指标选用 AIC (Akaike, 1981)、BIC (Schwarz, 1978)、CAIC (Bozdogan, 1987)和 SABIC (Sclove, 1987)，指标值相对越小表示模型对数据的拟合相对越好。另外，使用卡方检验评估 GDINA 与各约束模型之间对数据的拟合是否有显著差异。模型-数据绝对拟合指标选用 SRMSR (Maydeu-Olivares, 2013)[4]，Hu 和 Bentler (1999)指出 SRMSR 值小于 0.10 表示模型拟合数据。另外，使用题目区分度指标(IDI)来评估题目质量；使用分类精度指标(Wang et al., 2015)来计算分类信度。

另外，为探究在关键行动编码中引入迷思概念的相对优势，本研究还参考 Zhan 和 Qiao (2022)中不包含迷思概念的关键行动编码方式对该数据进行了编码。该数据包含 4 个问题解决技能(a1~a4)和 10 道虚拟题目，相应的 N-gram 字节片段序列和 Q 矩阵见附录 S3。

3 结果

表 4 呈现了两种编码方式下 4 个模型对数据的绝对和相对拟合结果。首先，由于两种编码方式得到的数据不同，在模型-数据拟合指标上两者没有可对比性。其次，针对引入迷思概念的编码方式获得的数据而言，仅有 GDINA 模型拟合该数据，且 AIC 和 SABIC 倾向于选择 GDINA 模型；尽管 BIC 和 CAIC 倾向于选择 DINA 模型，但 DINA 模型不拟合该数据且与饱和模型有显著差异。因此，该编码方式下的后续分析使用 GDINA 模型。再次，针对仅包含问题解决技能的编码方式获得的数据而言，仅有 DINO 模型不拟合数据；另外，尽管所有相对拟合指标均倾向于选择 DINA 模型，但鉴于 DINA 模型与饱和模型没有显著差异且为保证与引入迷思概念的编码方式可比，该编码方式下的后续分析也使用 GDINA 模型。

表 5 呈现了两种编码方式下的分类信度。整体而言，无论是属性水平还是测验整体水平的分类信度

[4] 本研究中由于模型参数过多，GDINA 包无法计算 $RMSEA_2$ 指标；这在 GDINA 包使用指导中也有提及(Shi et al., 2021)。

都较高，分析结果可靠。相对而言，额外引入迷思概念不仅没有降低分类信度，还使分类信度略有提高。

图 2 呈现了引入迷思概念的编码方式下的虚拟题目参数估计值(具体估计结果详见附录表 S2.1)。整体而言，除虚拟题目 AG、GC、ABH、AGC 和 FBC 外，其他虚拟题目都具有较高的题目质量。影响这 5 道题目质量的主要原因是它们的失误参数较高。结合图 2 和表 3 可发现，要呈现这 5 道虚拟题目所对应的行动序列，需要同时掌握问题解决技能和迷思概念，且在最优行动序列所含问题状态(e.g., A、B 和 C)和非最优行动序列所含问题状态(e.g., F、G 和 H)之间进行转换。因此，这 5 道虚拟题目的失误参数估计值较高意味着即便被试同时掌握题目所需的问题解决技能和迷思概念，也有较高概率不呈现这种跨最优-非最优问题状态的转移。另外，仅包含问题解决技能的关键行动编码方式下的虚拟题目参数估计值见附录表 S2.2。

表 4. 模型-数据拟合结果.

编码方式	模型	参数数量	SRMSR	−2LL	AIC	BIC	CAIC	SABIC	Δ*df*	*p*-value
a + m	GDINA	399	0.0756	27724.68	28522.69	30986.06	31385.06	29718.24		
	DINA	311	0.1006	27901.90	28523.91	30443.98	30754.98	29455.78	88	<0.001
	DINO	311	0.2483	58379.36	59001.36	60921.43	61232.43	59933.23	88	<0.001
	ACDM	342	0.1554	40126.82	40810.83	42922.29	43264.29	41835.59	57	<0.001
a	GDINA	67	0.0246	12753.24	12887.24	13300.88	13367.88	13087.99		
	DINA	35	0.0243	12786.02	12856.02	13072.10	13107.10	12960.89	32	0.43
	DINO	35	0.1316	18773.68	18843.67	19059.76	19094.76	18948.55	32	<0.001
	ACDM	45	0.0498	14918.08	15008.08	15285.90	15330.90	15142.91	22	<0.001

注: a + m = 包含问题解决技能和迷思概念的关键行动编码；a = 仅包含问题解决技能的关键行动编码.

表 5. 基于 GDINA 模型的分类信度结果.

编码方式	a1	a2	a3	a4	m1	m2	m3	m4	测验水平
a + m	0.9992	0.9998	0.9999	0.9998	0.9987	1.0000	1.0000	1.0000	0.9977
a	0.9936	0.9965	0.9938	0.9999					0.9858

注: a + m = 包含问题解决技能和迷思概念的关键行动编码；a = 仅包含问题解决技能的关键行动编码.

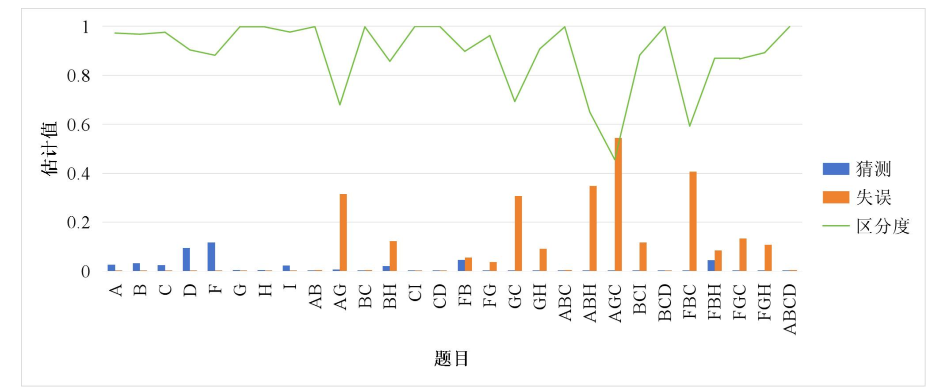


图 2. 基于 GDINA 模型的虚拟题目参数估计结果(已转换为猜测和失误参数). 注: 区分度 = 1−猜测−失误.

图 3 呈现了引入迷思概念的编码方式下各属性之间以及各属性与原始得分之间的多分格相关(polychoric correlation)矩阵(仅包含问题解决技能的编码方式的相关结果见附录图 S2.2)。首先，4 个问题解决技能之间呈中到高程度的正相关，且它们与原始得分之间成高程度正相关。其次，除 m4 和 m3 之

间成低程度负相关外，4 个迷思概念之间也呈中到高程度正相关，且它们与原始得分之间成中到高程度负相关。再次，除 m4 和 a3 之间成低程度正相关外，问题解决技能与迷思概念之间多为中到高程度负相关。上述三组相关关系中，m4 和 m3 以及 m4 和 a3 之间的相关系数正负略显特殊；结合图 2 和表 3，属性 a3、m3 和 m4 分别对应着“Individual”、“Daily”和“Other trips”。m4 和 m3 之间负相关表明被试在解决问题过程中并不倾向于同时呈现“Daily”和“Other trips”这两个操作。在该任务中，“Other trips”处于正确路径(i.e., 最优行动序列)上，而“Daily”处于错误路径(i.e., 非最优行动序列)上；被试只能通过“CANCEL”返回起始状态并做一系列操作后才能同时呈现这两个操作。付颜斌 等(2023)和 Han et al. (2022)的研究均发现，当被试已经处于正确或错误路径上的问题状态时，其更倾向于维持在该正确或错误路径上直到完成作答。另外，m4 和 a3 之间的正相关表明被试在解决问题过程中倾向于同时呈现“Individual”和“Other trips”；这是因为“Other trips”是“Individual”的下位操作。

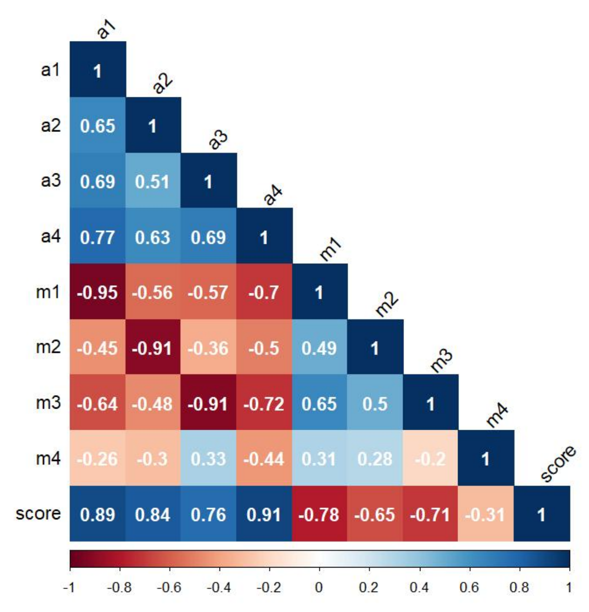


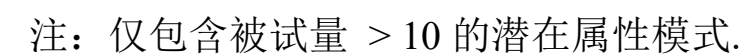
图 3. 属性之间以及属性与原始得分之间的相关(引入迷思概念的关键行动编码方式).

表 6 呈现了引入迷思概念的编码方式下被归入各潜在属性模式的被试量及其原始作答结果得分。在原始作答结果得 1 分的被试中，出现频率最高的潜在属性模式为(11110000)；而在原始作答结果得 0 分的被试中，出现频率最高的两个潜在属性模式为(01001010)和(00001110)。上述结果符合本研究逻辑，即掌握所有问题解决技能且不受迷思概念影响的被试，理论上可呈现最优行动序列并获得 1 分；而没有掌握所有问题解决技能且受迷思概念影响的被试，理论上难以呈现最优行动序列，进而有更大的可能性获得 0 分。另外，聚焦于原始作答结果得 1 分的被试，可发现一些被试在掌握 4 个问题解决技能的同时也掌握了个别迷思概念。比如，对于属性模式为(11110100)的被试，回顾其问题解决行动序列可发现该类被试在优惠类型选择环节(图 1(b))会在全价票和打折票之间反复操作、犹豫不决。图 4 进一步呈现了

两种编码方式下被试诊断分类结果的对应关系(各模式下具体人数见附录表 S2.3)。结果表明额外引入迷思概念可进一步实现对被试的精细化分类，更有益于探查导致其问题解决成败的具体原因。

表 6. 不同作答结果包含的潜在属性模式(引入迷思概念的关键行动编码方式).

作答结果得分	潜在属性模式	被试量
0	01001010	269
0	00001110	227
0	11000010	158
0	10000110	78
0	11100000	77
0	01111000	63
0	00101101	61
0	10110100	61
0	01101001	50
0	11100001	45
0	00101100	40
0	01101000	32
0	11000110	26
0	10100101	24
0	00111100	23
0	10100100	22
0	11001010	16
0	01001110	13
1	11110000	1681
1	11110100	78
1	11110110	74
1	11111000	63
1	11111010	56
1	11110010	43
1	11100001	40
1	11111100	16
1	11111101	13
1	11111110	12


注：仅包含被试量 ＞10 的潜在属性模式.

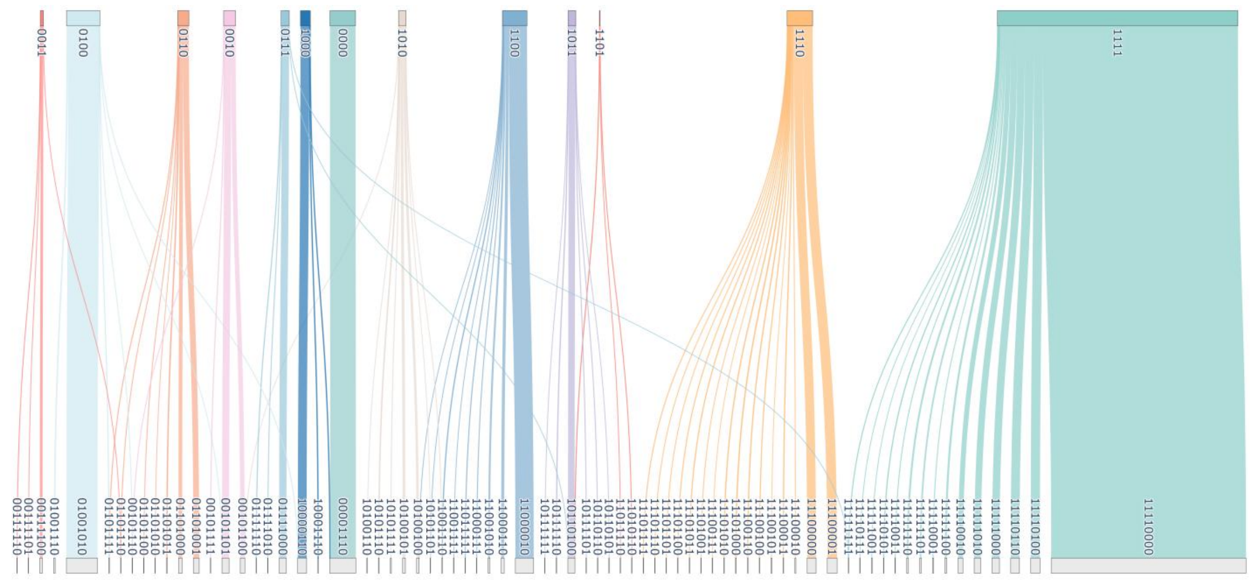


图 4. 两种关键行动编码方式下被试诊断分类结果对应图.
注: 4 属性模式为仅包含问题解决技能的编码方式下的诊断分类结果，8 属性模式为同时包含问题解决技能和迷思概念的编码方式下的诊断分类结果; 连接线粗细表示被试量.

4 总结与讨论

与传统作答精度等结果数据相比，行动序列等过程数据能提供更多有关被试如何解决问题的信息。但行动序列的非标准化格式也导致传统心理测量学模型无法被直接应用。目前，关键行动编码是将传统心理测量模型迁移应用至分析行动序列等过程数据的必要手段。为利用过程数据实现对问题解决技能的诊断分析，Zhan 和 Qiao (2022)通过将问题解决技能与关键行动序列相链接，提出了基于虚拟题目拓展的关键行动编码。然而，该编码方式忽略了迷思概念对问题解决过程的影响。对此，本研究提出一种引入迷思概念的关键行动编码方式，同时考虑了问题解决技能和迷思概念对问题解决过程的影响。使用一道来自 PISA 2012 的问题解决任务，对比探究了不额外引入迷思概念的和额外引入迷思概念的关键行为编码方式的实践表现。研究主要发现，同时考虑问题解决技能和迷思概念可实现对被试群体的更精细化分类，更有助于识别导致问题解决成败的具体原因，为个性化辅助教学提供更具针对性的补救建议。另外，还发现(1)额外引入迷思概念会略微提高诊断分类信度；(2)被试对迷思概念的掌握程度与原始作答结果得分之间呈中到高程度负相关。

需要提及的是，尽管本研究在解构行动序列时引入了迷思概念，但在分析中仍使用了常规 DCM 而非包含迷思概念的 DCM (e.g., Kuo et al., 2018; Ma et al., 2024)。其原因在于，本研究中潜在属性(迷思概念和问题解决技能)对外显行为(错误状态转移和正确状态转移)的影响机理是一致的(i.e., 假设掌握迷思概念导致错误状态转移与假设掌握问题解决技能导致正确状态转移的逻辑是一致的)，符合常规 DCM 的理论假设；而现有包含迷思概念的 DCM 中问题解决技能和迷思概念对外显行为的影响机理是不一致的。

当然，由于能力和精力有限，本研究仍有一些局限有待未来做进一步研究。第一，与 Zhan 和 Qiao (2022)等大多数行动序列分析方法研究类似，新编码方式仅考虑的行动序列本身所提供的信息，忽略了行动序列对应的行动时间(action time)中所包含的信息(e.g., Fu et al., 2023)；未来可尝试把行动时间视为虚拟题目作答时间，并引入题目作答时间模型或联合模型(e.g., van der Linden, 2007; Zhan et al., 2018)进行分析。第二，与绝大多数行动序列分析方法研究一样，本研究仅聚焦于一道问题解决题目；理论上，只要多道问题解决题目测量相同的属性，新编码方式也适用于一次性解构多道题目的行动序列并进行诊断分类分析。第三，本研究聚焦于任务目标明确且已知信息完备的结构良好(well-defined)问题解决任务；这类任务通常拥有相对明确且数量有限的问题状态(i.e., 有明确的状态转移规则)。然而，现实中还有很多具有不良结构的问题解决任务，缺乏明确的状态转移规则；如何对这类任务实施关键行动编码仍是一个有待解决的问题。第四，本研究编码过程中将多次重复行动(e.g., 重复点击某一个按钮)压缩为一次重复行动，忽略了重复行动中有可能反映被试因属性掌握熟练度不够或认知风格导致的犹豫不决的潜在测量信息；未来研究可尝试探究如何更好地利用这些信息，实现对个体问题解决素养更全面的理解。

总之，本研究提出了一种引入迷思概念的关键行动编码方式并探究了其在过程数据诊断分类分析中的应用。新编码方式的提出有助于研究者进一步识别影响被试问题解决成败的具体原因，对实施有针对

性干预提供方法学支持：为提高被试的问题解决表现，除尝试提高其问题解决技能外，还可考虑消除迷思概念的负面影响。

参考文献

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, *16*(1), 3–14. https://doi.org/10.1016/0304-4076(81)90071-3

Bergner, Y., & von Davier, A. A. (2019). Process Data in NAEP: Past, Present, and Future. *Journal of Educational and Behavioral Statistics*, *44*(6), 706–732.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370.https://doi.org/10.1007/BF02294361

Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: A psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika*, *79*(3), 403–425. https://doi.org/10.1007/s11336-013-9350-4

Chen, F., Lu, C., & Cui, Y. (2023). Using learners' problem-solving processes in computer-based assessments for enhanced learner modeling: A deep learning approach. *Education and Information Technologies.* https://doi.org/10.1007/s10639-023-12389-x

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

Ding, S. L., Yang, S. Q., & Wang, W. Y. (2010). The importance of reachability matrix in constructing cognitively diagnostic testing. *Journal of Jiangxi Normal University (National Sciences Education), 34*(5), 490–494. https://doi.org/10.16357/j.cnki.issn1000-5862.2010.05.023

[丁树良，杨淑群，汪文义. (2010). 可达矩阵在认知诊断测验编制中的重要作用. *江西师范大学学报(自然科学版), 34*(05), 490–494.]

Eggen, P., & Kauchak, D. (2004) *Educational psychology: Windows, classrooms*. Upper Saddle River: Pearson Prentice Hall.

Fu Y., Chen Q., & Zhan P. (2023). Binary modeling of action sequences in problem-solving tasks: One- and two-parameter action sequence model. *Acta Psychologica Sinica, 55*(8), 1383–1404. https://doi.org/10.3724/SP.J.1041.2023.01383

[付颜斌，陈琦鹏，詹沛达.(2023).问题解决任务中行动序列的二分类建模：单/两参数行动序列模型. *心理学报*, *55*(08), 1383–1404.]

Fu, Y., Zhan, P., Chen, Q., & Jiao, H. (2023). Joint modeling of action sequences and action time in computer-based interactive tasks. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02178-2

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior*, *61*, 36–46. https://doi.org/10.1016/j.chb.2016.02.095

Gu, Y., & Xu, G. (2020). Partial identifiability of restricted latent class models. *The Annals of Statistics, 48*(4), 2082–2107. https://www.jstor.org/stable/26931550

Han, Y., Liu, H., & Ji, F. (2022). A Sequential Response Model for Analyzing Process Data on Technology-Based Problem-Solving Tasks. *Multivariate Behavioral Research*, *57*(6), 960–977. https://doi.org/10.1080/00273171.2021.1932403

Hao, J., & Shu, Z. (2015). Analyzing process data from game/scenario- based tasks: An edit distance approach. *Journal of Educational Data Mining, 7*(1), 33–50.

Hao, J., Smith, L., Mislevy, R., Von Davier, A., & Bauer, M. (2016). Taming log files from game/simulation-based assessments: Data models and data analysis tools. *ETS Research Report Series*, *2016*(1), 1–17. https://doi.org/10.1002/ets2.12096

He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, *166*, 104170. https://doi.org/10.1016/j.compedu.2021.104170

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258–272. https://doi.org/10.1177/01466210122032064

Kuo, B.-C., Chen, C.-H., & De La Torre, J. (2018). A cognitive diagnosis model for identifying coexisting skills and misconceptions. *Applied Psychological Measurement*, *42*(3), 179–191. https://doi.org/10.1177/0146621617722791

Kuo, B.-C., Chen, C.-H., Yang, C.-W., & Mok, M. M. C. (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items. *Educational Psychology*, *36*(6), 1115–1133. https://doi.org/10.1080/01443410.2016.1166176

Levy, R. (2019). Dynamic Bayesian network modeling of game-based diagnostic assessments. *Multivariate Behavioral Research*, *54*(6), 771–794. https://doi.org/10.1080/00273171.2019.1590794

Li, M., Liu, H., Cai, M., & Yuan, J. (2023). Estimation of individuals' collaborative problem solving ability in computerbased assessment. *Education and Information Technologies, 29*, 483–515. https://doi.org/10.1007/s10639-023-12271-w

Li, M., Liu, Y.,& Liu H.(2020).Analyzing problem-solving strategies in computer-adaptive testing: An expansion and application of multi-level hybrid IRT models. *Acta Psychologica Sinica, 52*(4), 528–540. https://doi.org/10.3724/SP.J.1041.2020.00528

[李美娟, 刘玥, 刘红云 (2020). 计算机动态测验中问题解决过程策略的分析：多水平混合 IRT 模型的拓展与应用. *心理学报*, *52*(4), 528–540.]

Liu, H., Han, Y., Xiao, Y., Yuan, J., & Li, M. (2023). Evaluation and development of process-oriented measurement models for complex problem-solving skills. *China Examinations, 11*, 9–20. https://doi.org/10.19360/j.cnki.11-3303/g4.2023.11.002.

[刘红云，韩雨婷，肖悦，袁建林，李美娟. (2023). 复杂问题解决能力的过程性测评与测量模型发展.*中国考试,11*, 9–20.]

Liu, H., Liu, Y., & Li, M. (2018). Analysis of process data of PISA 2012 computer-based problem solving: Application of the modified multilevel mixture IRT model. *Frontiers in Psychology, 9,* 1372. https://doi.org/10.3389/fpsyg.2018.01372

Liu, Y., Xu, H., Chen, Q., & Zhan, P. (2022). The measurement of problem-solving competence using process data. *Advances in Psychological Science, 30*(3), 522–535. https://doi.org/10.3724/SP.J.1042.2022.00522

[刘耀辉，徐慧颖，陈琦鹏，詹沛达. (2022). 基于过程数据的问题解决能力测量及数据分析方法.*心理科学进展, 30*(3), 522–538.]

Ma, W., Sorrel, M. A., Zhai, X., & Ge, Y. (2024). A dual-purpose model for binary data: Estimating ability and misconceptions. *Journal of Educational Measurement*. Advance online publication. https://doi.org/10.1111/jedm.12383

Ma, W., & de la Torre, J. (2020). GDINA: An *R* package for cognitive diagnosis modeling. *Journal of Statistical Software*, *93*(14). https://doi.org/10.18637/jss.v093.i14

Martin, R. E., Sexton, C. M., & Gerlovich, J. A. (2001). Teaching science for all children. Allyn and Bacon.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement, 11*(3), 71–101.

Mislevy, R. J. (2019). Advances in measurement and cognition. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 164–182. https://doi.org/10.1177/0002716219843816

OECD.(2014).*PISA 2012 results：Creative problem solving: Students' skills in tackling real-life problems (Volume V)*.PISA,OECD Publishing:https://dio.org/10.1787/9789264208070-en

Qiao, X., & Jiao, H. (2018). Data mining techniques in analyzing process data: A didactic. *Frontiers in Psychology*, *9*, 2231. https://doi.org/10.3389/fpsyg.2018.02231

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2).461-464. https://doi.org/10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333–343. https://doi.org/10.1007/BF02294360

Shi, Q., Ma, W., Robitzsch, A., Sorrel, M. A., & Man, K. (2021). Cognitively diagnostic analysis using the G-DINA model in R. *Psych, 3*(4), 812–835. https://doi.org/10.3390/psych3040052

Sun, X., Li, J., & Fu, Z. (2018). Predicting students' reasoning ability and math grades using game log-file - Application of machine learning. *Acta Psychologica Sinica, 50*(7), 761–770. https://doi.org/10.3724/SP.J.1041.2018.00761

[孙鑫，黎坚，符植煜. (2018). 利用游戏 log-file 预测学生推理能力和数学成绩——机器学习的应用. *心理学报*, *50*(7), 761–770.]

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*(3), 287–308. https://doi.org/10.1007/s11336-006-1478-z

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*(3), 456–477. https://doi.org/10.1111/bmsp.12054

Xia, S., Zhan, P., Chan, K. K. H., & Wang, L. (2023). Assessing concept mapping competence using item expansion-based diagnostic classification analysis. *Journal of Research in Science Teaching*. Advance online publication. https://doi.org/10.1002/tea.21897

Xiao, Y., & Liu, H. (2023). A state response measurement model for problem-solving process data. *Behavior Research Methods, 56*(1), 258–277. https://doi.org/10.3758/s13428-022-02042-9

Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, *71*(2), 262–286. https://doi.org/10.1111/bmsp.12114

Zhan, P., & Qiao, X. (2022). Diagnostic classification analysis of problem-solving competence using process data: An item expansion method. *Psychometrika*, *87*(4), 1529–1547. https://doi.org/10.1007/s11336-022-09855-9

Key Action Encoding Incorporating Misconceptions and Its Application in Diagnostic Classification Analysis of Process Data

Fangfang Gao Qipeng Chen Peida Zhan

(School of Psychology, Zhejiang Normal University)

Abstract

Process data encompasses the human-computer interaction data captured in computer-based learning and assessment systems, reflecting participants’ problem-solving processes. Among various types of process data, action sequences stand out as a quintessential type, delineating participants’ step-by-step problem-solving processes. However, the non-standardized format of action sequences, characterized by varying data lengths among participants, presents challenges for the direct application of traditional psychometric models like diagnostic classification models (DCM). Extending psychometric models applicable to standardized structured data to process data analysis often necessitates key-action encoding – determining if each participant's data contains essential problem-solving actions and encoding them (e.g., “1” for "contains" and “0” for "does not contain"). Zhan and Qiao (2022) proposed a key-action encoding method facilitating the application of DCM to process data analysis for identifying participants’ mastery of problem-solving skills. Nevertheless, their approach overlooks the adverse impact of misconceptions on problem-solving. To this end, this study

introduces a key-action encoding approach incorporating misconceptions and explores its utility in diagnostic classification analysis of process data. This new encoding method integrates both problem-solving skills and misconceptions, extending Zhan and Qiao's (2022) approach.

An illustrative example is provided to compare the performance of the proposed encoding approach with Zhan and Qiao's (2022) approach using a real-world interactive assessment item, "Tickets," from PISA 2012. For the proposed approach, eight attributes (four problem-solving skills and four misconceptions) and 28 phantom items (i.e., key actions) were defined based on the scoring rule and assessment framework of the interactive assessment item. In contrast, Zhan and Qiao's approach defined four attributes (problem-solving skills) and 10 phantom items. Four DCMs – DINA, DINO, ACDM, and GDINA models – were employed for data analysis. The relative fit metrics for model-data comparison were selected from AIC, BIC, CAIC, and SABIC. Additionally, a chi-square test was employed to evaluate whether there existed a significant difference in the fit to the data between GDINA and each of the constrained models. For assessing absolute fit between the model and the data, the SRMSR metric was utilized. Moreover, item quality was evaluated using the item differentiation index (IDI), while classification reliability was determined by calculating the classification accuracy index.

The findings reveal that: (1) considering both problem-solving skills and misconceptions enables more nuanced participant classification, facilitating identification of specific factors influencing problem-solving success and failure and offering targeted remedial suggestions for personalized instruction; (2) the introduction of misconceptions slightly enhances diagnostic classification reliability; (3) a moderate-to-high negative correlation exists between participants' mastery of misconceptions and raw scores, indicating misconceptions diminish students' overall problem-solving performance.

In summary, this study proposes a key-action encoding approach incorporating misconceptions and explores its application in diagnostic classification analysis of process data, specifically action sequences. The proposed approach aids researchers in pinpointing specific factors influencing problem-solving outcomes and provides methodological support for targeted interventions. To enhance participants' problem-solving performance, beyond improving their skills, addressing misconceptions' adverse effects merits consideration.

Keywords cognitive diagnosis; process data; problem solving; misconception; action sequence

附　录

S1. 过程数据预处理

S1.1. 过程数据整理

为便于研究，我们依据如下步骤对原始log-file数据进行预处理。以图S1为例，首先，由于event_value列为后四列的汇总(network、fare_type、ticket_type和number_trips)，所以删除后四列数据；其次，将前三列(cnt、schoolid和StIDStd)合并至最后一列，并命名为compounds_id；最后，根据compounds_id，对所有被试排序(字母由前到后，数字由小到大)，形成初始过程数据。

cnt	schoolid	StIDStd	event	time	event_number	event_value	network	fare_type	ticket_type	number_trips
AUS	0000032	00618	START_ITEM	0.1000	1.00	NULL	NULL	NULL	NULL	NULL
AUS	0000032	00618	ACER_EVENT	50.8000	2.00	country_trains	country_trains	NULL	NULL	0
AUS	0000032	00618	ACER_EVENT	53.0000	3.00	full_fare	country_trains	full_fare	NULL	0
AUS	0000032	00618	ACER_EVENT	58.9000	4.00	individual	country_trains	full_fare	individual	0
AUS	0000032	00618	ACER_EVENT	66.3000	5.00	trip_2	country_trains	full_fare	individual	2
AUS	0000032	00618	ACER_EVENT	77.2000	6.00	Buy	country_trains	full_fare	individual	2
AUS	0000032	00618	END_ITEM	80.9000	7.00	NULL	NULL	NULL	NULL	NULL

(a) 原始log-file数据

cnt	schoolid	StIDStd	event	time	event_number	event_value
AUS	0000032	00618	START_ITEM	0.1000	1.00	NULL
AUS	0000032	00618	ACER_EVENT	50.8000	2.00	country_trains
AUS	0000032	00618	ACER_EVENT	53.0000	3.00	full_fare
AUS	0000032	00618	ACER_EVENT	58.9000	4.00	individual
AUS	0000032	00618	ACER_EVENT	66.3000	5.00	trip_2
AUS	0000032	00618	ACER_EVENT	77.2000	6.00	Buy
AUS	0000032	00618	END ITEM	80.9000	7.00	NULL

(b) 删除后四列

cnt	schoolid	StIDStd	event	time	event_number	event_value	compounds_id
AUS	0000032	00618	START_ITEM	0.1000	1.00	NULL	AUS000003200618
AUS	0000032	00618	ACER_EVENT	50.8000	2.00	country_trains	AUS000003200618
AUS	0000032	00618	ACER_EVENT	53.0000	3.00	full_fare	AUS000003200618
AUS	0000032	00618	ACER_EVENT	58.9000	4.00	individual	AUS000003200618
AUS	0000032	00618	ACER_EVENT	66.3000	5.00	trip_2	AUS000003200618
AUS	0000032	00618	ACER_EVENT	77.2000	6.00	Buy	AUS000003200618
AUS	0000032	00618	END ITEM	80.9000	7.00	NULL	AUS000003200618

(c) 前三列合并为“compounds_id”

图S1.1. PISA2012 TICKETS task1(CP038Q02)问题解决测验中某被试的log-file数据预处理过程

S1.2. 过程数据编码

AUS	0000369	06813	START_ITEM	0.1000	1.00	NULL	AUS000036906813
AUS	0000369	06813	ACER_EVENT	37.6000	2.00	F	AUS000036906813
AUS	0000369	06813	ACER_EVENT	42.6000	3.00	B	AUS000036906813
AUS	0000369	06813	ACER_EVENT	48.7000	4.00	Cancel	AUS000036906813
AUS	0000369	06813	ACER_EVENT	51.9000	5.00	A	AUS000036906813
AUS	0000369	06813	ACER_EVENT	52.6000	6.00	B	AUS000036906813
AUS	0000369	06813	ACER_EVENT	55.7000	7.00	C	AUS000036906813
AUS	0000369	06813	ACER_EVENT	56.9000	8.00	D	AUS000036906813
AUS	0000369	06813	ACER_EVENT	79.4000	9.00	D	AUS000036906813
AUS	0000369	06813	ACER_EVENT	80.5000	10.00	Buy	AUS000036906813
AUS	0000369	06813	END ITEM	84.2000	11.00	NULL	AUS000036906813

(a) 初始过程数据

AUS	0000369	06813	ACER_EVENT	37.6000	2.00	F	AUS000036906813
AUS	0000369	06813	ACER_EVENT	42.6000	3.00	B	AUS000036906813
AUS	0000369	06813	ACER_EVENT	51.9000	5.00	A	AUS000036906813
AUS	0000369	06813	ACER_EVENT	52.6000	6.00	B	AUS000036906813
AUS	0000369	06813	ACER_EVENT	55.7000	7.00	C	AUS000036906813
AUS	0000369	06813	ACER_EVENT	56.9000	8.00	D	AUS000036906813
AUS	0000369	06813	ACER_EVENT	79.4000	9.00	D	AUS000036906813

(b) 删除起始状态(“START_ITEM”)、目标状态(“BUY”)和“Cancel”

AUS	0000369	06813	ACER_EVENT	37.6000	2.00 F	AUS000036906813
AUS	0000369	06813	ACER_EVENT	42.6000	3.00 B	AUS000036906813
AUS	0000369	06813	ACER_EVENT	51.9000	5.00 A	AUS000036906813
AUS	0000369	06813	ACER_EVENT	52.6000	6.00 B	AUS000036906813
AUS	0000369	06813	ACER_EVENT	55.7000	7.00 C	AUS000036906813
AUS	0000369	06813	ACER_EVENT	56.9000	8.00 D	AUS000036906813

(c) 删除相邻的重复问题状态

图 S1.2. 过程数据编码示例

S2. 补充结果

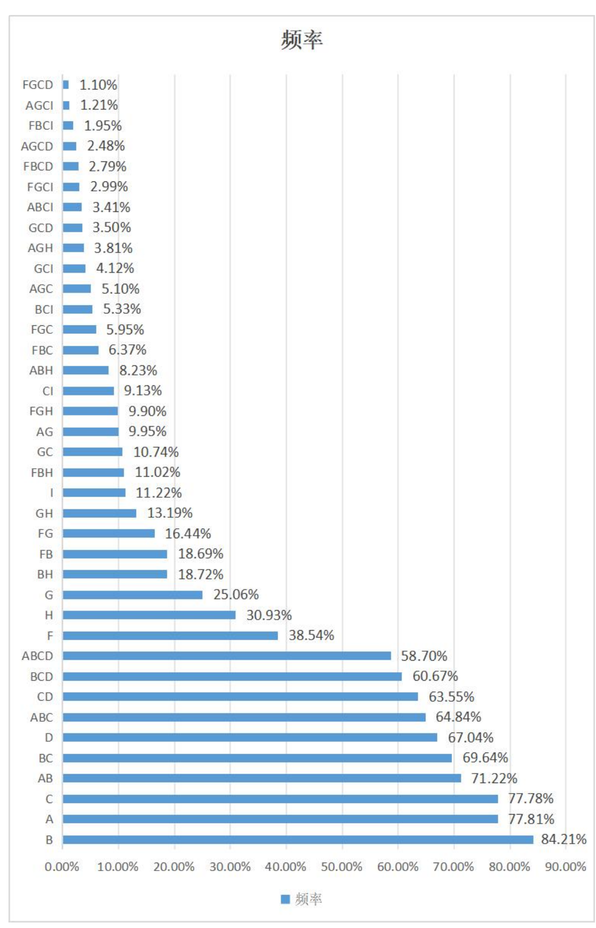


图 S2.1. 所有行动序列的呈现频率(引入迷思概念的关键行动编码方式).

表 S2.1. 基于 GDINA 模型的虚拟题目参数估计值(引入迷思概念的关键行动编码方式).

题目	猜测（标准误)	失误（标准误)	区分度
A	0.0270 (0.0089)	0.0001 (0.2736)	0.9729
B	0.0316 (0.1229)	0.0001 (0.2935)	0.9683
C	0.0237 (0.0951)	0.0001 (0.4354)	0.9762
D	0.0960 (0.0116)	0.0001 (1.1998)	0.9039
F	0.1177 (0.0084)	0.0001 (0.3540)	0.8822
G	0.0011 (0.0413)	0.0001 (0.2371)	0.9988
H	0.0020 (0.0303)	0.0001 (0.1812)	0.9979

I	0.0230 (0.0029)	0.0001 (0.1793)	0.9769
AB	0.0001 (0.0062)	0.0008 (0.0044)	0.9991
AG	0.0068 (0.2562)	0.3135 (0.0844)	0.6797
BC	0.0001 (0.3558)	0.0016 (0.0014)	0.9983
BH	0.0205 (0.2252)	0.1220 (0.0555)	0.8575
CI	0.0001 (0.4430)	0.0001 (0.0872)	0.9998
CD	0.0001 (1.5468)	0.0001 (0.5725)	0.9998
FB	0.0467 (0.1784)	0.0551 (0.0270)	0.8982
FG	0.0001 (0.4816)	0.0372 (0.0224)	0.9627
GC	0.0001 (0.1458)	0.3067 (0.1118)	0.6932
GH	0.0001 (0.2621)	0.0916 (0.0319)	0.9083
ABC	0.0001 (0.1837)	0.0016 (0.0012)	0.9983
ABH	0.0001 (0.3344)	0.3495 (0.0979)	0.6504
AGC	0.0001 (0.3095)	0.5440 (0.1245)	0.4559
BCI	0.0001 (0.3453)	0.1168 (0.0342)	0.8831
BCD	0.0001 (0.3453)	0.0001 (0.0185)	0.9998
FBC	0.0001 (0.4907)	0.4071 (0.0485)	0.5928
FBH	0.0450 (0.3163)	0.0850 (0.0329)	0.8700
FGC	0.0001 (0.3655)	0.1326 (0.0616)	0.8673
FGH	0.0001 (0.2271)	0.1071 (0.0387)	0.8928
ABCD	0.0001 (0.1837)	0.0005 (0.0008)	0.9994

表 S2.2. 基于 GDINA 模型的虚拟题目参数估计值(仅包含问题解决技能的关键行动编码方式).

题目	猜测（标准误)	失误（标准误)	区分度
A	0.0491 (0.0241)	0.0001 (0.0926)	0.9508
B	0.0318 (0.0164)	0.0001 (0.5879)	0.9681
C	0.0580 (0.0174)	0.0001 (0.2547)	0.9419
D	0.0975 (0.0104)	0.0001 (0.3744)	0.9024
AB	0.0001 (0.3238)	0.0001 (0.0432)	0.9998
BC	0.0001 (0.4284)	0.0001 (0.1085)	0.9998
CD	0.0001 (0.6648)	0.0001 (0.0336)	0.9998
ABC	0.0001 (0.3064)	0.0011 (0.0235)	0.9988
BCD	0.0001 (0.1509)	0.0001 (1.2563)	0.9998
ABCD	0.0001 (0.3064)	0.0001 (0.4702)	0.9998

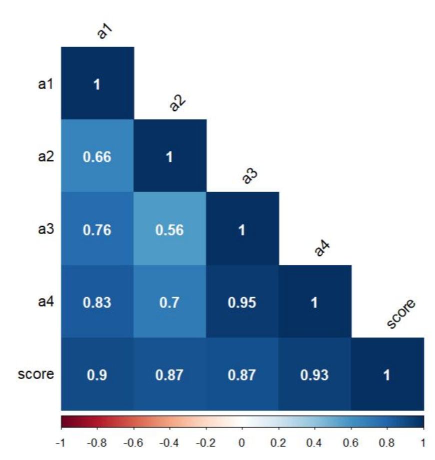


图 S2.2. 属性之间以及属性与原始得分之间的相关(仅包含问题解决技能的关键行动编码方式).

表 S2.3. 两种关键行动编码方式下被试诊断分类结果对应表(仅呈现被试量大于 10 的属性模式).

作答结果得分	仅包含问题解决技能	同时包含问题解决技能和迷思概念	被试量
1	1111	11110000	1681
		11110100	78
		11110110	74
		11111000	63
		11111010	56
		11110010	43
		11111100	16
		11111101	13
		11111110	12
1	1110	11100001	40
0	0000	00001110	222
0	0100	01001010	269
		01001110	13
0	1100	11000010	158
		11000110	26
		11001010	16
0	1110	11100000	77
		11100001	45
0	0010	00101101	60
		00101100	39
0	0110	01101001	50
		01101000	32
0	1000	10000110	76
0	0111	01111000	63
0	1011	10110100	57
0	1010	10100101	24
		10100100	21
0	0011	00111100	23

S3. 仅包含问题解决技能(不包含迷思概念)的关键行动编码

表 S3.1 呈现了该任务下从一元到四元字节片段序列对应的行动序列。此时，可以将被试所呈现的任一行动序列视为表中字节片段(或组合)。表 S3.2 呈现了包含 4 个问题解决技能和 10 道虚拟题目的 Q 矩阵(10 道虚拟题目对应的行动序列的呈现频率均处于 5%~95%之间，未删题)。

表 S3.1. PISA2012 问题解决测验 TICKETS (CP038Q02)任务中 N-gram 字节片段序列(不含迷思概念).

N-gram	数量	字节片段序列
Uni-gram	4	A、B、C、D
Bi-gram	3	AB、BC、CD
Tri-gram	2	ABC、BCD
Quad-gram	1	ABCD

表 S3.2. PISA2012 问题解决测验 TICKETS (CP038Q02)任务的 Q 矩阵(不含迷思概念).

题号	题目	问题解决技能			
		a1	a2	a3	a4
1	A	1	0	0	0
2	B	0	1	0	0
3	C	0	0	1	0
4	D	0	0	0	1
5	AB	1	1	0	0
6	BC	0	1	1	0
7	CD	0	0	1	1
8	ABC	1	1	1	0
9	BCD	0	1	1	1
10	ABCD	1	1	1	1